

Projektnummer: B1 Betong

Tom Pedersen, Natanael Blomberg & Piere Ventura Cruz

2021-10-10

1 Sammanfattning

I det här projektet analyserar vi data som kommer från en undersökning av statens provningsanstalt. Från experimentet gavs 4 mått (ID-min, ID-max, YD-min, YD-max) som alla innehåller 10 kolumner som representerar varje rör och 11 rader där rad 1 är huvudanstalten medan raderna 2-11 är för varje filial. Vi utför sedan test inom och mellan måtten. Där vi genom anovatabeller kom fram till att F-testet för filialerna förkastar H_0 : alla filialer mäter likvärdigt. F-testet för mått förkastar också H_0 : de olika måtten mäts likvärdigt.

Men vi kunde verkligen inte förkasta att samspel inte finns. Vi utförde även grafer i form av boxplottar, normalfördelningsplottar och interactionplottar för att kunna stärka vår tes ytterligare och inte bara lita helt blint på värdena från anovatabellen. Där kom vi fram till att våra tidigare antaganden stämmer och att vi med hjälp utav detta kunde beräkna hur mycket ett genomsnittligt rör avviker från perfekt rundhet. Vi kollade medelvärden mellan måtten och inom filialerna för att se hur mycket som skiljer från huvudanstaltens resultat. Vi testade med alla filialer men även utav filial 4 och 6 då det såg ut att vara en väldigt tvivelaktig observation. Men det slutliga resultatet var att filial 4 och 6 inte gjorde någon skillnad utan resultatet var ungefär densamma om inte något sämre.

2 Inledning

I det här projektet kommer vi att använda oss utav variansanalys för att studera given data. Vi har ett dataset med information från en undersökning av statens provningsanstalt gällande kontrollmätning utav betongrör.

Statens provningsanstalt har filialer ute i landet som utför provningar och mätningar. Dessa kontrolleras emellanåt utav huvudanstalten. Vid en sådan kontroll sändes 10 betongrör runt till 10 olika filialer för en mätning utav 4 diametermått per rör.

Vår uppgift blir då att hjälpa provningsanstalten med att analysera datat men särskilt att undersöka huruvida filialerna mäter likvärdigt, om de olika måtten mäts likvärdigt vid de olika mätlokalerna men även hur mycket som avviker ett genomsnittligt rör från perfekt rundhet i respektive ände.

3 Data

Vår data innehåller 4 olika min- och maxmått som motsvarar den inre och yttre diametrarna på betongrören som skall undersökas. Varje mått består utav 10 kolumner och 11 rader. Varje kolumn svarar mot ett visst betongrör och för var och en av de fyra diametrarna utgörs rad 1 av mätdata från huvudanstalten medan rad 2 till 11 av mätdata från filialerna. För en fullständig överblick av data hänvisas till appendix.

Nedan följer en sammanfattning av variationsområdet för data. Vi kan genast observera att rör 1 har ett största mätvärde som avviker kraftigt från övriga, vilket korrigeras från 7777 till 2777.

rör 1	rör 2	rör 3	rör 4	rör 5
Min. :2771	Min. :2758	Min. :2761	Min. :2787	Min. :2777
1st Qu.:2809	1st Qu.:2817	1st Qu.:2794	1st Qu.:2796	1st Qu.:2787
Median :2928	Median :2888	Median :2872	Median :2866	Median :2862
Mean :2980	Mean :2871	Mean :2864	Mean :2867	Mean :2863
3rd Qu.:2939	3rd Qu.:2938	3rd Qu.:2938	3rd Qu.:2938	3rd Qu.:2938
Max. :7777	Max. :2946	Max. :2945	Max. :2943	Max. :2945
rör 6	rör 7	rör 8	rör 9	rör 10
Min. :2780	Min. :2776	Min. :2771	Min. :2764	Min. :2746
1st Qu.:2791	1st Qu.:2786	1st Qu.:2787	1st Qu.:2794	1st Qu.:2818
Median :2862	Median :2862	Median :2862	Median :2872	Median :2890
Mean :2863	Mean :2862	Mean :2861	Mean :2864	Mean :2869
3rd Qu.:2935	3rd Qu.:2938	3rd Qu.:2937	3rd Qu.:2938	3rd Qu.:2938
Max. :2949	Max. :2946	Max. :2944	Max. :2943	Max. :2943

4 Statistisk modellering

Med hjälp av vår data utförs statistisk modellering för att komma fram till olika resultat för att besvara våra frågeställningar. Där vi främst väljer att använda oss utav ANOVA-tabeller där vi för multipel regression beräknar ett F-värde som sedan kommer att hjälpa oss besvara våra frågeställningar. Om p -värdet för respektive F-testet är mindre än vår signifikansnivå kan detta användas för att förkasta nollhypotesen och tvärtom om det skulle visa sig att F-testet ej visar sig vara signifikant. För att utföra den statistiska analysen börjar vi med att subtrahera huvudanstaltens mätningar från respektive filial, så att vi får deras relativa mätfel. För en inblick av den data som används för den statistiska analysen hänvisas till appendix, *formaterad data*.

De första frågorna vi vill besvara är om de tio filialerna mäter likvärdig och om det är någon eller några som sticker ut, samt om de fyra diametermåten mäts likvärdigt. För detta ändamål användes följande modell

$$mätfel_{ijk} = \mu + filial_i + mått_j + (filial \times mått)_{ij} + rör_k + \varepsilon_{ijk} \quad (1)$$

där $\varepsilon_{ijk} \sim N(0, \sigma^2)$ och oberoende feltermar och μ är medelvärdet av $mätfel$ för all data. I modell (1) behandlas $filial$ och $mått$ samt deras samspel $filial \times mått$ som systematiska faktorer eftersom frågorna berör huruvida dessa har någon effekt på mätfelet. Faktorn $rör$ behandlas som slumpmässig eftersom det inte finns något som tyder på att statens provningsanstalt har valt dessa systematiskt och eftersom rören anses vara av samma typ med samma nominella dimensioner.

Ett viktigt bivillkor för en meningsfull statistisk analys under modell (1) är bivillkoret

$$\sum_i filial_i = \sum_j mått_j = \sum_{ij} (filial \times mått)_{ij} = 0,$$

det vill säga att nivåerna för respektive systematisk faktor summerar till 0.

Nollhypotesen för respektive systematisk faktor är följande:

$$filial : H_0: filial_1 = filial_2 = \dots = filial_{10} = 0;$$

$$mått: H_0: mått_1 = mått_2 = mått_3 = mått_4 = 0;$$

$$filial \times mått: H_0: (filial \times mått)_{1,1} = \dots = (filial \times mått)_{10,4} = 0,$$

där samspelet $filial \times mått$ antas vara lika för alla möjliga nivåkombinationer.

Resultatet av modell (1. 1): *all data* presenteras i nedan sammanfattning av ANOVA-tabellen:

```
Error: rör
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  91.47   10.16

Error: Within
      Df Sum Sq Mean Sq F value Pr(>F)
filial    9  1350  150.05   43.49 <2e-16 ***
mått     3   415  138.34   40.10 <2e-16 ***
filial:mått 27  2472   91.57   26.54 <2e-16 ***
Residuals 351  1211    3.45

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tabell 1.1: Uförande av F-test i en variansanalystabell. Modell; $mätfel \sim filial + mått + filial:mått + Error(rör)$.

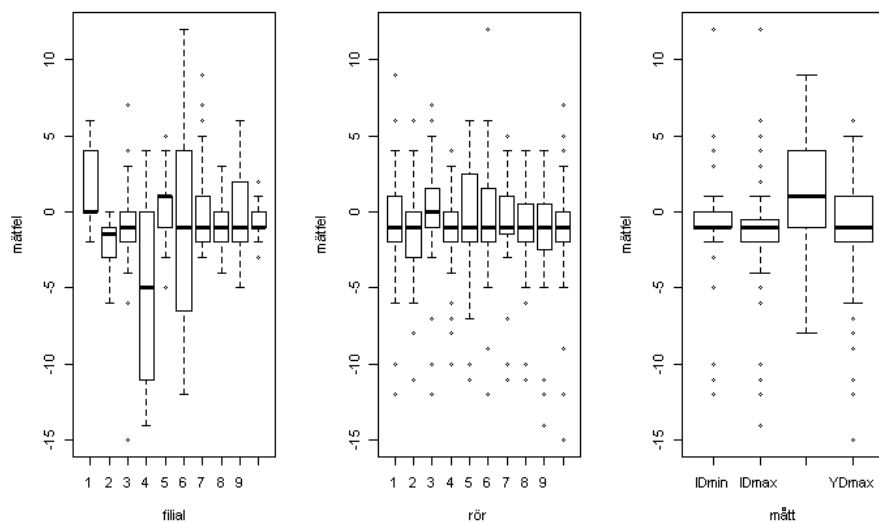
I **Tabell 1.1** ser vi att mätfelet anges som en modell av faktorerna filial, mått och filial:mått, men också att data ger stöd för att mätfelet skiljer sig mellan filialer samt för de olika diametermått. Detta märker vi genom att vi fått så extremt låga p-värden som i sin tur innebär att vi kan förkasta nollhypotesen för respektive test på 95%-nivån. Vi kan alltså förkasta hypotesen att filialerna mäter lika och att diametermått mätas på ett likvärdigt sätt. Dessutom kan vi avläsa att det finns ett signifikant samspel mellan filialerna och de olika diametermått vilket antyder att filialerna inte mäter dessa på ett likvärdigt sätt.

Tables of effects

filial										
filial	1	2	3	4	5	6	7	8	9	10
	2.352	-1.197	-0.223	-4.747	1.452	-0.098	1.002	0.102	0.952	0.402
mått										
mått	IDmin	IDmax	YDmin	YDmax						
	-0.5375	-0.8675	1.7325	-0.3275						
filial:mått										
	mått									
filial	IDmin	IDmax	YDmin	YDmax						
1	-1.262	-1.233	0.567	1.928						
2	1.588	1.817	-3.083	-0.322						
3	0.813	0.343	0.843	-1.997						
4	-4.762	-4.233	3.967	5.027						
5	-0.063	0.468	-0.132	-0.272						
6	4.987	5.718	-3.783	-6.923						
7	-0.913	-1.282	2.318	-0.122						
8	-0.013	-0.583	-0.182	0.778						
9	-0.763	-1.332	1.168	0.927						
10	0.387	0.317	-1.682	0.978						

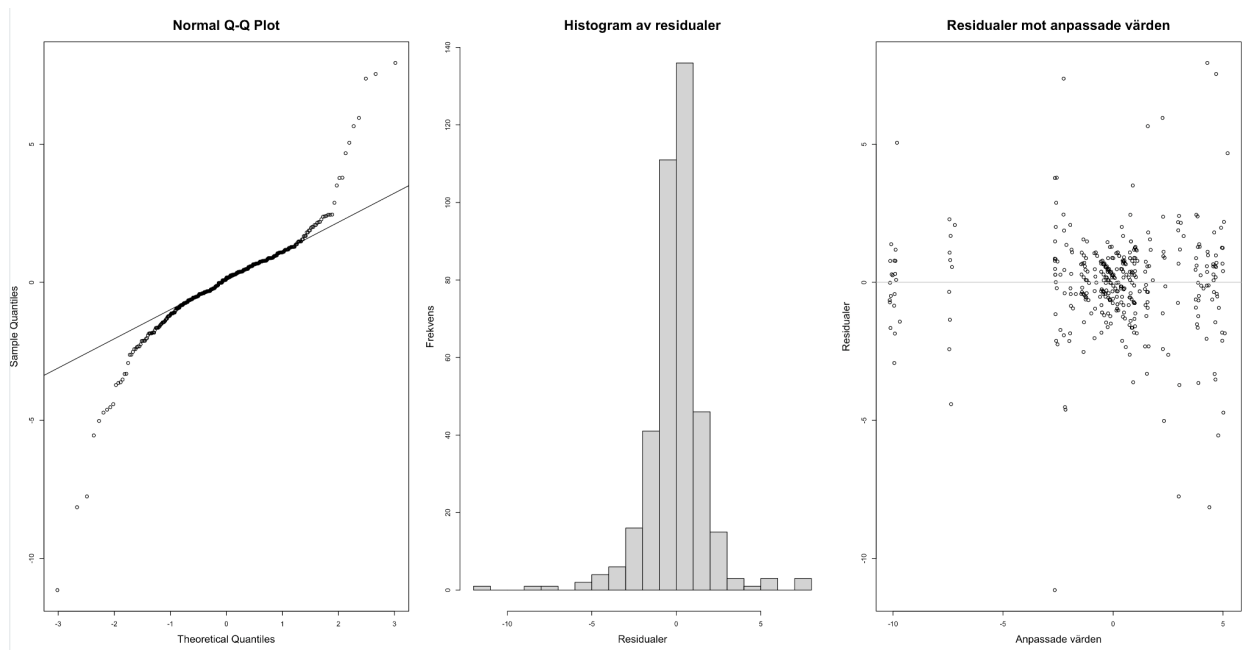
Tabell 1.2: Illustration av avvikelser (nivåeffekter) från det vägda genomsnittet μ för faktorerna filial, mått och filial:mått.

I Tabell 1.2 noterar vi att nivåeffekterna för samspelet för filial 4 respektive filial 6 avviker kraftigt från de övriga. Det finns alltså anledning att tro att filial 4 respektive filial 6 avviker från övriga filialer, åtminstone för någon av måtten. Vi kommer även att illustrera plottar som kommer att visas nedanför.



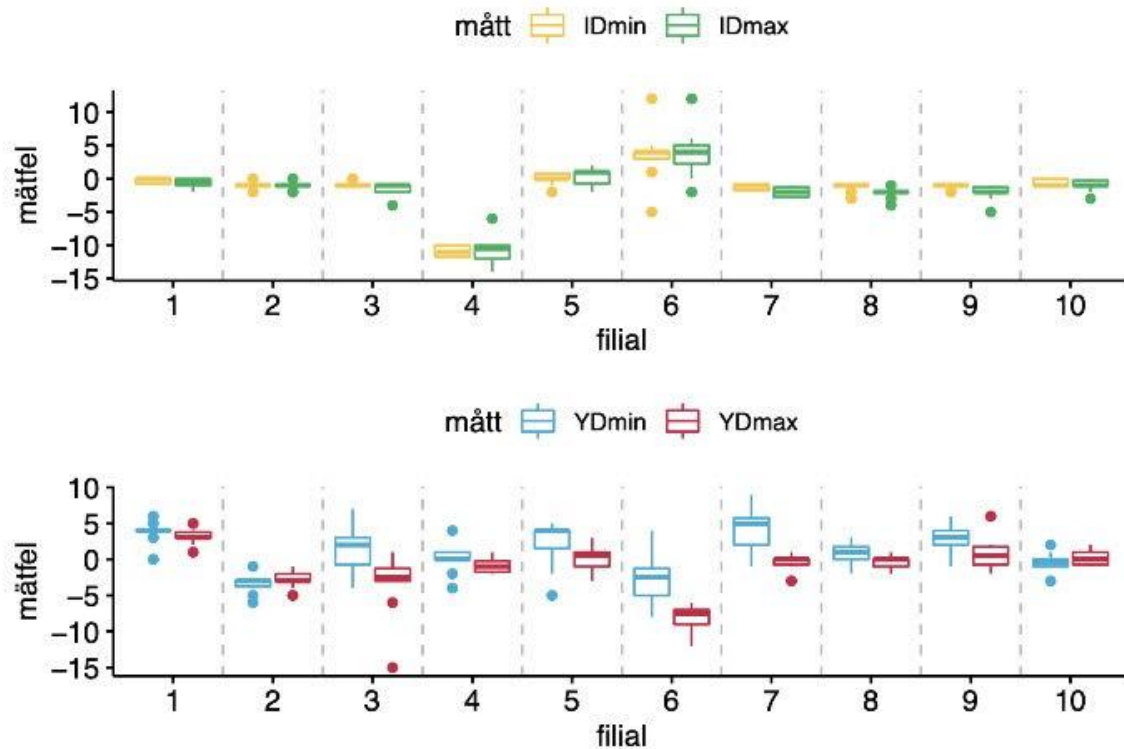
Figur 1: Boxplottarna illustreras i följande ordning; mätfel gentemot filial, rör och mått.

I Figur 1 ser vi en väldigt stor spridning för filial 4 och filial 6 medan resterande filialer har en median för mätfelen relativt nära noll. Väljer vi att titta på dem olika måtten så ser vi att ID har ett betydligt bättre resultat än YD, även om outliers förekommer för både ID och YD, då medianerna och kvartilerna inte sammanfaller lika väl för YD. Vi kan även notera att det inte verkar finnas någon större skillnad mellan de olika rören, vilket motiverar att vi kan definiera dessa som en slumpmässig faktor i modell (1.1).



Figur 2: Q-Q residualer, histogram av residualerna och residualer mot anpassade värden.

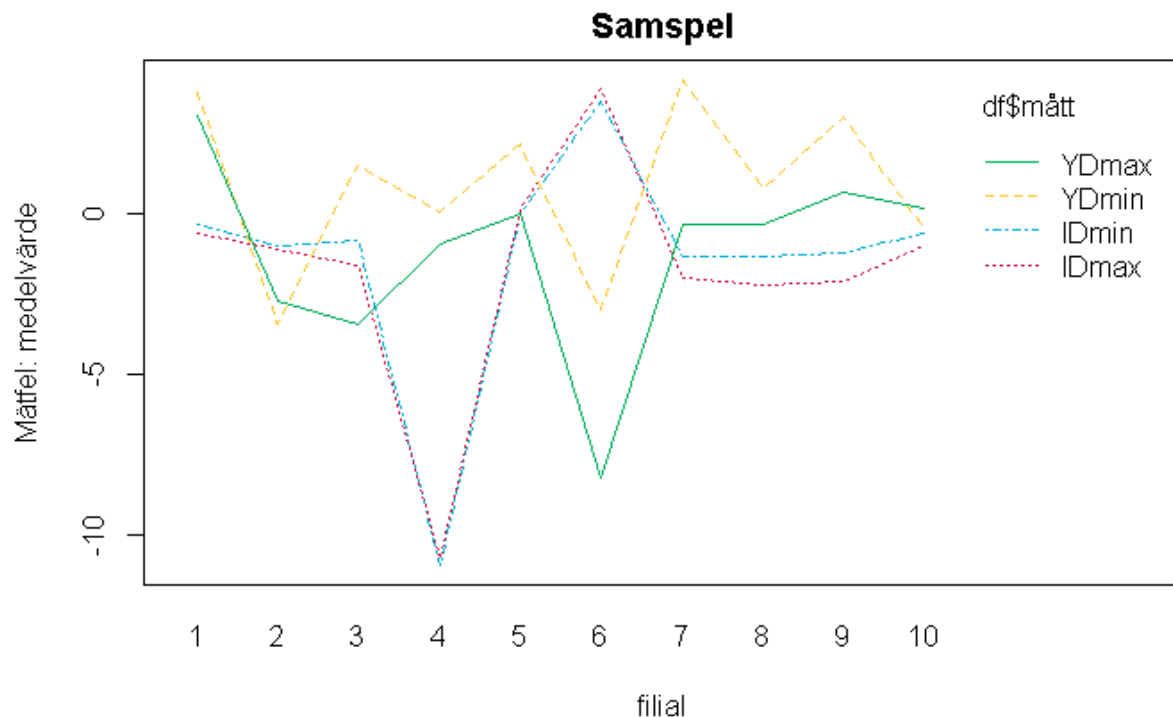
Vi observerar från *Figur 2* att QQ-plotten ser ut att ha problem med svansarna, vilket antyder att vi har relativt få men samtidigt en ansevärd mängd punkter som ligger för långt ifrån den centrerade massan för att data skulle komma från en normalfördelning. Dock ser fördelningen fortfarande ut att vara symmetrisk. En anledning till detta skulle kunna vara vi har data som kommer från två eller flera fördelningar med ungefär samma väntevärde men väldigt olika varians. Om vi betraktar residualer mot anpassade värden ser vi att väntevärdet för residualerna ser ut att vara ungefär noll, men att det verkar finnas en systematisk skillnad i variansen.



Figur 3: Boxplottar över mätfelen ID min/max (övre) respektive YD min/max (undre).

Vi observerar från Figur 3 att de flesta värdena när det kommer till ID ser bra ut men att just filial 4 och 6 sticker ut kraftigt. För resterande filialer (alla utom 4 och 6) ser vi att variationen både mellan och inom grupper ser ut att vara relativt liten.

För YD så ser vi att det är större spridning både mellan och inom grupper. För ID var det endast två filialer som hade en markant avvikelse från 0 medan det ser ut att vara betydligt fler när det kommer till YD. Något man däremot ska notera är att YD max och YD min inte avviker lika kraftigt utan att alla håller sig någorlunda nära 0. Vi noterar även att filial 4 inte har samma relativa mätfel för YD som för ID.



Figur 4: Samspel mellan filial och diametermått mot deras medelvärde.

I Figur 4 ser vi ett väldigt litet samspel inom måtten för innerdiameter och ett visst samspel inom måtten för ytterdiameter. Däremot ser vi ett väldigt stort samspel mellan måtten för inner och ytterdiameter. Vi kan alltså ana att skillnaden mellan hur filialerna mäter diametermått främst är skillnaden mellan de två måtten för ID mot de två måtten för YD, men att det även finns ett visst samspel mellan max- och minmått inom YD. Vi kan även se att filial 4 avviker kraftigt från resterande filialer inom ID-måtten samt att filial 6 avviker både inom ID- och YD-måtten.

Eftersom Figur 4 visade på att skillnaden i hur filialerna mäter de olika diametermått främst beror på skillnader mellan ID- och YD-måtten väljer vi att testa modell (1) då vi skiljer på dessa. Vi får alltså samma modell men med olika delmängder av data och följaktligen något annorlunda hypoteser.

Undersökning av innerdiametermått (ID)

Vi börjar med att undersöka om de tio filialerna mäter likvärdig och om det är någon eller några som sticker ut, samt om de två diametermått mäts likvärdigt, då vi endast betraktar innerdiametermått. Modell (1. 2) som används är samma som modell (1) med skillnaden att index j endast antar värden 1 och 2, dvs vi betraktar endast måtten IDmin och IDmax. Resultat av denna modell presenteras i appendix under *modell (1. 2)*. Anledningen är att vi efter den grafiska diagnostiken fortfarande inte kunde motivera att residualerna var normalfördelade med konstant varians. Av denna anledning, med motivation Figur 3: *Boxplottar över mätfelen*, väljer vi att återigen testa modell (1) men då vi endast har med IDmin och IDmax bland diametermått, samt att vi utesluter filial 4 och filial 6 för att se om vi kan få en modell som bättre uppfyller modellantagandena. Detta eftersom filial 4 och 6 avviker mycket från övriga filialer.

Resultatet av modell (1. 3) för ID, filial 4 och 6 exkluderade, presenteras i nedan sammanfattning av ANOVA-tabellen.

```
Error: rör
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  21.81    2.423

Error: Within
      Df Sum Sq Mean Sq F value    Pr(>F)
filial    7  59.54   8.506  16.914 6.49e-16 ***
mått      1   9.51   9.506  18.902 2.69e-05 ***
filial:mått  7   5.74   0.821   1.632  0.132
Residuals 135  67.89   0.503

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tabell 2.1: Uförande av F-test i en variansanalystabell. Modell; mätfel ~ filial + mått + filial:mått + Error(rör). YDmin, YDmax samt filial 4 och 6 exkluderade.

Om vi betraktar *Tabell 2.1* ser vi att det är en signifikant skillnad mellan IDmin och YDmin men att det inte finns någon signifikant skillnad i hur de olika filialerna mäter IDmin och IDmax, ty F-testet för samspelet har p -värdet 0.132. För denna modell finns det fortfarande ett litet problem med att QQ-plotten (appendix: modell (1. 3)) har en vänstersvans, men i residualplotten finns inte ett lika stort problem med en systematisk skillnad i varians. Därmed är modellantagande bättre uppfyllda för denna modell än de föregående vilket gör F-testet mer tillförlitligt.

Slutsatsen är att även om vi exkluderar de starkt avvikande filialerna 4 och 6 så finns det fortfarande en signifikant skillnad mellan filialerna och mellan IDmin och IDmax, men det finns ingen signifikant skillnad i hur filialerna mäter IDmin och IDmax.

Undersökning av ytterdiametern (YD)

Vidare undersöka vi om de tio filialerna mäter likvärdig och om det är någon eller några som sticker ut, samt om de två diametermått mätts likvärdigt, då vi endast betraktar ytterdiametermått. Modellen som används är samma som modell (1) med skillnaden att index j endast antar värden 3 och 4, dvs vi betraktar endast måtten YDmin och YDmax.

Resultatet av modell (1. 4) för YD presenteras i nedan sammanfattning av ANOVA-tabellen.


```

Error: rör
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  110.9   12.32

Error: Within
      Df Sum Sq Mean Sq F value    Pr(>F)
filial    9 1238.2   137.58  30.415 < 2e-16 ***
mått      1   212.2    212.18  46.907 1.28e-10 ***
filial:mått  9   212.7    23.64   5.225 2.81e-06 ***
Residuals 171   773.5     4.52
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Tabell 3.1: Uförande av F-test i en variansanalystabell. Modell; mätfel ~ filial + mått + filial:mått + Error(rör). IDmin och IDmax exkluderade.

I *tabell 3.1* ser vi att det finns en signifikant skillnad mellan filialerna, mellan YDmin och YDmax samt hur de olika filialerna mäter YDmin och YDmax. Med avseende på *Figur 6* i appendix (modell (1. 4)) ser vi att QQ-plotten har problem med svansarna och att residualplotten har ett litet problem med systematisk varians. Även här ser det ut som om variationen mellan grupper inte är likvärdig. I detta fall är det svårt att se om det är någon eller några filialer som avviker kraftigt från de övriga och vi väljer att förlita oss på ovan modell för att dra slutsatser rörande frågeställningen med avseende på YD-måtten.

För att besvara frågan om rörens rundhet vill vi finna filialer som mäter likvärdigt med avseende på min- och maxmått. Vi vill alltså finna filialer som har liten variation mellan YDmin och YDmax, samt liten variation inom dessa mätningar. För detta ändamål väljer vi att testa modell (1) för YD med filial 2, 8 och 10, eftersom dessa verkar uppfyller ovan krav med avseende på *Figur 3: Boxplottar över mätfelen*.

Resultatet av modell (1. 5) för YD med filial 2, 8 och 10 presenteras i nedan sammanfattning av ANOVA-tabellen.

```

Error: rör
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9   33.6    3.733

Error: Within
      Df Sum Sq Mean Sq F value    Pr(>F)
filial    2  131.43    65.72   60.849 1.6e-13 ***
mått      1    0.07     0.07    0.062  0.8049
filial:mått  2   10.23     5.12    4.738  0.0136 *
Residuals  45   48.60     1.08
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Tabell 4.1: Uförande av F-test i en variansanalystabell. Modell; mätfel ~ filial + mått + filial:mått + Error(rör). Endast YD-mått och filial 2, 8 och 10.

I *Tabell 4.1* ser vi att det inte finns någon signifikant skillnad mellan YDmin och YDmax, vilket var målet med ovan modell. I appendix (modell (1.5)) ser vi att både QQ-ploten och residualploten antyder att vi har en modell som uppfyller modellantagandena väl.

Rörens rundhet

För att kvantifiera hur mycket ett genomsnittligt rör avviker från perfekt rundhet i respektive ände jämför vi dess min- och maxmått för ID respektive YD. Vi gör detta genom att beräkna medelvärde för den procentuella skillnaden mellan min- och maxmåttens dels för huvudanstalten, samtliga filialer samt ett urval av filialerna baserat på ovan modeller.

En genomsnittlig procentuell avvikelse A från perfekt rundhet beräknas enligt nedan

$$A = \left[\frac{1}{10} \sum_{k=1}^{10} \frac{1}{|F|} \sum_{i \in F} \frac{(D_{max} - D_{min})_{k,i}}{(D_{max} + D_{min})_{k,i}/2} \right] \cdot 100 \%$$

där k är summation över alla rör, F är indexmängden för de utvalda filialerna och D antingen är ID eller YD. I fallet då vi använder huvudanstaltens mätningar summerar vi endast över dessa. För ID valde vi att använda alla filialer eftersom motsvarande modell (1.2) hade p -värde för faktorn *mått* som var 0.109. Vi testade även att exkludera filial 4 och 6 eftersom dessa avvek så mycket från övriga filialer, vilket motsvarar modell (1.3). För YD valde vi modell (1.5) som har ett p -värde på 0.805 för faktorn *mått*.

5 Resultat

Den generella modellen vi tycker besvarar frågeställningen bäst blir

$$mätfel_{ijk} = \mu + filial_i + mått_j + (filial \times mått)_{ij} + rör_k + \varepsilon_{ijk} \quad (1)$$

där $\varepsilon_{ijk} \sim N(0, \sigma^2)$ är oberoende feltermen och μ är medelvärdet av $mätfel$ för all data. För modell (1) behandlas $filial$ och $mått$ samt deras samspel $filial \times mått$ som systematiska faktorer eftersom frågorna berör huruvida dessa har någon effekt på mätfelet. Faktorn $rör$ behandlas som slumpmässig eftersom det inte finns något som tyder på att statens provningsanstalt har valt dessa systematiskt och eftersom rören anses vara av samma typ med samma nominella dimensioner.

För att besvara samtliga frågor användes modell (1) men med olika delmängder av data. En sammanfattning följer nedan, tillsammans med svar av frågeställningarna under inledningen;

- Mäter de olika filialerna lika bra?
- Är det någon som klart avviker från de övriga?
- Mäts de fyra diametermått på ett likvärdigt sätt vid de olika mätlokalerna?

Utifrån modell (1. 1), som använder all data, kunde vi se att det fanns en signifikant skillnad i hur de olika filialerna mäter, samt att de finns en signifikant skillnad i hur de olika mätlokalerna mäter de fyra diametermått. Från modellen kunde vi även se att nivåeffekterna för filial 4 och filial 6 avvek kraftigt. Ett problem med modellen var att den grafiska diagnostiken inte motiverade att residualerna var normalfördelade med konstant varians.

I modell (1. 3) undersökte vi endast måtten för innerdiameter för alla filialer exklusive 4 och 6, som avvek kraftigt. Resultat var att det fortfarande fanns en signifikant skillnad mellan filialerna och innerdiameterens min- och maxmått, men det fanns ingen signifikant skillnad i hur filialerna mäter det minsta och största måtten för innerdiameteren. För denna modell motiverade den grafiska diagnostiken att residualerna i högre grad har konstant varians och väntevärde noll, men inte nödvändigtvis att dessa var normalfördelade. Dock var modellantagandena i högre grad uppfyllda för denna modell kontra (1. 1).

I modell (1. 4) undersökte vi endast måtten för rörens ytterdiameter. Resultatet blev att det fanns en signifikant skillnad mellan filialerna, mellan största och minsta mått för ytterdiameteren, samt hur respektive filial mäter minsta och största värde. Även i denna modell motiverade den grafiska diagnostiken att residualerna har konstant varians och väntevärde noll, men inte nödvändigtvis att dessa var normalfördelade.

Vidare skulle vi besvara följande fråga:

- Hur mycket avviker ett genomsnittligt rör från perfekt rundhet i respektive ände?

För att undersöka rörens rundhet söktes en modell där det minsta och största diametermättet mättes på ett likvärdigt sätt. I detta avseende borde modell (1. 2) vara den bästa för innerdiameter, med ett p -värde på 0.109 för faktorn *mått*. Vi valde även att testa modell (1. 3) där filial 4 och 6 är exkluderade.

För att testa ytterdiameter konstruerade vi en ny modell utifrån *Figur 3: Boxplottar över mätfelen*, som endast använder måtten från filial 2, 8 och 10. Modellen har ett p -värde på 0.805 för faktorn *mått*. Resultatet presenteras nedan.

Avvikelse från perfekt rundhet

ID	Huvudanstalt	Alla filialer	Filial 4 och 6 exkluderade
Avvikelse %	0.129	0.118	0.113

YD	Huvudanstalt	Alla filialer	Filial 2, 8 och 10
Avvikelse %	1.352	1.278	1.355

I ovan tabell ser vi att de modeller som inte visade någon signifikant skillnad mellan hur filialerna mäter min- och maxmått för inner- och ytterdiameter presterade bäst relativt huvudanstalten.

För ID kunde vi genom beräkning av medelvärden se att de skiljer sig en endast 0.011%-enheter från perfekt rundhet. Vi testade även att göra oss av med filial 4 och 6 som kunde ge ett missvisande resultat men kom fram till att detta inte var fallet utan att skillnaden till och med ökade en aning. När det kommer till YD så ser vi att filialens mått avviker med 0.064%-enheter från perfekt rundhet.

6 Diskussion

Alltså har vi kommit fram till att filialerna inte mäter likvärdigt. Detta då att vi genom anova tabellen kunde förkasta vårt H_0 : alla filialer mäter likvärdigt. Vi har på liknande vis kommit fram till att F-testet för mått också förkastar H_0 : de olika måtten mäts likvärdigt.

Under den statistiska modelleringen hade vi vid flera tillfällen problem med att residualerna inte var normalfördelade. För att undvika detta problem definierade vi flera olika modeller som innehöll färre nivåer av de systematiska faktorerna. En annorlunda, och eventuellt mer effektiv metod, hade kunnat vara att undersöka någon annan statistisk metod eller testat att transformera data.

Vidare hade en förbättring varit att undersöka residualerna för den slumpmässiga faktorn *rör*.

I den statistiska analysen har vi utgått från att huvudanstalten mäter korrekt, och i den meningen hade vi kunnat använda dess mätvärden för att direkt uttala oss om rörens rundhet. Vi valde dock att försöka uttala oss om rörens rundhet utifrån mått av anstaltens filialer för att få en frågeställning av mer statistisk betydelse.

7 Gruppmedlemmarnas arbetsinsats

All kod och beräkningar har i princip gjorts av alla för varje del. Detta för att säkerställa att det vi kommit fram till är något som vi alla är överens om och tycker är trovärdigt. Vi har alla varit med och tolkat resultatet och bidragit i analysen.

A Appendix: Grafer, R-kod, output om modeller

Data

ID MIN
2937 2938 2937 2938 2938 2935 2937 2937 2938 2938
2937 2938 2937 2938 2937 2935 2936 2936 2938 2938
2937 2937 2936 2937 2937 2934 2935 2936 2937 2937
2936 2937 2937 2937 2937 2934 2936 2937 2937 2937
2927 2927 2927 2928 2927 2923 2927 2926 2926 2926
2938 2938 2937 2939 2937 2935 2938 2937 2938 2936
2940 2933 2941 2942 2943 2947 2941 2941 2941 2939
2936 2937 2936 2937 2936 2934 2935 2936 2936 2937
2936 2937 2936 2937 2936 2934 2934 2936 2937 2937
2936 2937 2936 2937 2936 2934 2936 2936 2936 2937
2937 2938 2936 2938 2937 2934 2936 2936 2937 2938

ID MAX
2942 2946 2941 2940 2940 2937 2941 2940 2943 2941
2941 2946 2941 2940 2939 2936 2941 2939 2941 2941
2941 2945 2941 2939 2939 2935 2940 2939 2941 2940
2940 2945 2940 2939 2939 2935 2940 2938 2939 2940
2930 2940 2931 2930 2930 2925 2930 2930 2929 2929
2942 2944 2942 2941 2941 2939 2942 2941 2941 2940
2948 2944 2945 2943 2945 2949 2946 2944 2943 2943
2939 2943 2940 2939 2938 2935 2940 2938 2940 2939
2940 2942 2939 2938 2938 2935 2940 2938 2940 2939
2940 2943 2939 2938 2939 2935 2940 2938 2938 2940
2941 2944 2941 2939 2939 2936 2941 2939 2940 2941

YD MIN
2776 2761 2768 2795 2780 2785 2779 2776 2767 2749
2779 2765 2773 2795 2786 2789 2783 2780 2771 2753
2771 2758 2765 2789 2779 2781 2776 2773 2764 2746
2775 2762 2771 2791 2783 2788 2783 2775 2767 2756
2777 2761 2768 2791 2784 2786 2780 2776 2767 2747
2780 2765 2772 2793 2785 2780 2782 2780 2771 2750
2771 2759 2761 2787 2777 2783 2778 2771 2771 2748
2785 2763 2775 2794 2785 2791 2784 2778 2769 2754
2777 2763 2770 2793 2783 2786 2779 2776 2768 2749
2779 2760 2774 2796 2784 2791 2783 2778 2769 2752
2776 2760 2769 2792 2782 2785 2778 2776 2766 2748

YD MAX
2813 2842 2812 2800 2794 2797 2794 2795 2813 2854
2816 2845 2817 2804 2797 2800 2796 2798 2817 2855
2810 2838 2811 2798 2792 2794 2791 2792 2812 2849
2810 2836 2813 2799 2791 2795 2791 2793 2813 2839
2812 2841 2813 2799 2792 2797 2792 2794 2813 2852
2812 2840 2813 2801 2794 2800 2795 2796 2810 2853
2807 2834 2800 2793 2787 2788 2787 2789 2802 2845
2813 2839 2811 2799 2795 2798 2794 2795 2813 2854
2812 2842 2813 2800 2793 2797 2793 2795 2814 2852
2815 2848 2810 2799 2795 2799 2794 2796 2811 2854
2813 2843 2814 2801 2793 2797 2793 2795 2814 2853

Formaterad data

	mätfel	filial	mätt	rör	51	3	6 IDmin	1	100	0	10 IDmin	10	151	6	6 IDmax	1
1	0	1	IDmin	1	52	-5	6 IDmin	2	101	-1	1 IDmax	1	152	-2	6 IDmax	2
2	0	1	IDmin	2	53	4	6 IDmin	3	102	0	1 IDmax	2	153	4	6 IDmax	3
3	0	1	IDmin	3	54	4	6 IDmin	4	103	0	1 IDmax	3	154	3	6 IDmax	4
4	0	1	IDmin	4	55	5	6 IDmin	5	104	0	1 IDmax	4	155	5	6 IDmax	5
5	-1	1	IDmin	5	56	12	6 IDmin	6	105	-1	1 IDmax	5	156	12	6 IDmax	6
6	0	1	IDmin	6	57	4	6 IDmin	7	106	-1	1 IDmax	6	157	5	6 IDmax	7
7	-1	1	IDmin	7	58	4	6 IDmin	8	107	0	1 IDmax	7	158	4	6 IDmax	8
8	-1	1	IDmin	8	59	3	6 IDmin	9	108	-1	1 IDmax	8	159	0	6 IDmax	9
9	0	1	IDmin	9	60	1	6 IDmin	10	109	-2	1 IDmax	9	160	2	6 IDmax	10
10	0	1	IDmin	10	61	-1	7 IDmin	1	110	0	1 IDmax	10	161	-3	7 IDmax	1
11	0	2	IDmin	1	62	-1	7 IDmin	2	111	-1	2 IDmax	1	162	-3	7 IDmax	2
12	-1	2	IDmin	2	63	-1	7 IDmin	3	112	-1	2 IDmax	2	163	-1	7 IDmax	3
13	-1	2	IDmin	3	64	-1	7 IDmin	4	113	0	2 IDmax	3	164	-1	7 IDmax	4
14	-1	2	IDmin	4	65	-2	7 IDmin	5	114	-1	2 IDmax	4	165	-2	7 IDmax	5
15	-1	2	IDmin	5	66	-1	7 IDmin	6	115	-1	2 IDmax	5	166	-2	7 IDmax	6
16	-1	2	IDmin	6	67	-2	7 IDmin	7	116	-2	2 IDmax	6	167	-1	7 IDmax	7
17	-2	2	IDmin	7	68	-1	7 IDmin	8	117	-1	2 IDmax	7	168	-2	7 IDmax	8
18	-1	2	IDmin	8	69	-2	7 IDmin	9	118	-1	2 IDmax	8	169	-3	7 IDmax	9
19	-1	2	IDmin	9	70	-1	7 IDmin	10	119	-2	2 IDmax	9	170	-2	7 IDmax	10
20	-1	2	IDmin	10	71	-1	8 IDmin	1	120	-1	2 IDmax	10	171	-2	8 IDmax	1
21	-1	3	IDmin	1	72	-1	8 IDmin	2	121	-2	3 IDmax	1	172	-4	8 IDmax	2
22	-1	3	IDmin	2	73	-1	8 IDmin	3	122	-1	3 IDmax	2	173	-2	8 IDmax	3
23	0	3	IDmin	3	74	-1	8 IDmin	4	123	-1	3 IDmax	3	174	-2	8 IDmax	4
24	-1	3	IDmin	4	75	-2	8 IDmin	5	124	-1	3 IDmax	4	175	-2	8 IDmax	5
25	-1	3	IDmin	5	76	-1	8 IDmin	6	125	-1	3 IDmax	5	176	-2	8 IDmax	6
26	-1	3	IDmin	6	77	-3	8 IDmin	7	126	-2	3 IDmax	6	177	-1	8 IDmax	7
27	-1	3	IDmin	7	78	-1	8 IDmin	8	127	-1	3 IDmax	7	178	-2	8 IDmax	8
28	0	3	IDmin	8	79	-1	8 IDmin	9	128	-2	3 IDmax	8	179	-3	8 IDmax	9
29	-1	3	IDmin	9	80	-1	8 IDmin	10	129	-4	3 IDmax	9	180	-2	8 IDmax	10
30	-1	3	IDmin	10	81	-1	9 IDmin	1	130	-1	3 IDmax	10	181	-2	9 IDmax	1
31	-10	4	IDmin	1	82	-1	9 IDmin	2	131	-12	4 IDmax	1	182	-3	9 IDmax	2
32	-11	4	IDmin	2	83	-1	9 IDmin	3	132	-6	4 IDmax	2	183	-2	9 IDmax	3
33	-10	4	IDmin	3	84	-1	9 IDmin	4	133	-10	4 IDmax	3	184	-2	9 IDmax	4
34	-10	4	IDmin	4	85	-2	9 IDmin	5	134	-10	4 IDmax	4	185	-1	9 IDmax	5
35	-11	4	IDmin	5	86	-1	9 IDmin	6	135	-10	4 IDmax	5	186	-2	9 IDmax	6
36	-12	4	IDmin	6	87	-1	9 IDmin	7	136	-12	4 IDmax	6	187	-1	9 IDmax	7
37	-10	4	IDmin	7	88	-1	9 IDmin	8	137	-11	4 IDmax	7	188	-2	9 IDmax	8
38	-11	4	IDmin	8	89	-2	9 IDmin	9	138	-10	4 IDmax	8	189	-5	9 IDmax	9
39	-12	4	IDmin	9	90	-1	9 IDmin	10	139	-14	4 IDmax	9	190	-1	9 IDmax	10
40	-12	4	IDmin	10	91	0	10 IDmin	1	140	-12	4 IDmax	10	191	-1	10 IDmax	1
41	1	5	IDmin	1	92	0	10 IDmin	2	141	0	5 IDmax	1	192	-2	10 IDmax	2
42	0	5	IDmin	2	93	-1	10 IDmin	3	142	-2	5 IDmax	2	193	0	10 IDmax	3
43	0	5	IDmin	3	94	0	10 IDmin	4	143	1	5 IDmax	3	194	-1	10 IDmax	4
44	1	5	IDmin	4	95	-1	10 IDmin	5	144	1	5 IDmax	4	195	-1	10 IDmax	5
45	-1	5	IDmin	5	96	-1	10 IDmin	6	145	1	5 IDmax	5	196	-1	10 IDmax	6
46	0	5	IDmin	6	97	-1	10 IDmin	7	146	2	5 IDmax	6	197	0	10 IDmax	7
47	1	5	IDmin	7	98	-1	10 IDmin	8	147	1	5 IDmax	7	198	-1	10 IDmax	8
48	0	5	IDmin	8	99	-1	10 IDmin	9	148	1	5 IDmax	8	199	-3	10 IDmax	9
49	0	5	IDmin	9	100	0	10 IDmin	10	149	-2	5 IDmax	9	200	0	10 IDmax	10
50	-2	5	IDmin	10					150	-1	5 IDmax	10				

Modell 1.2

Error: rör

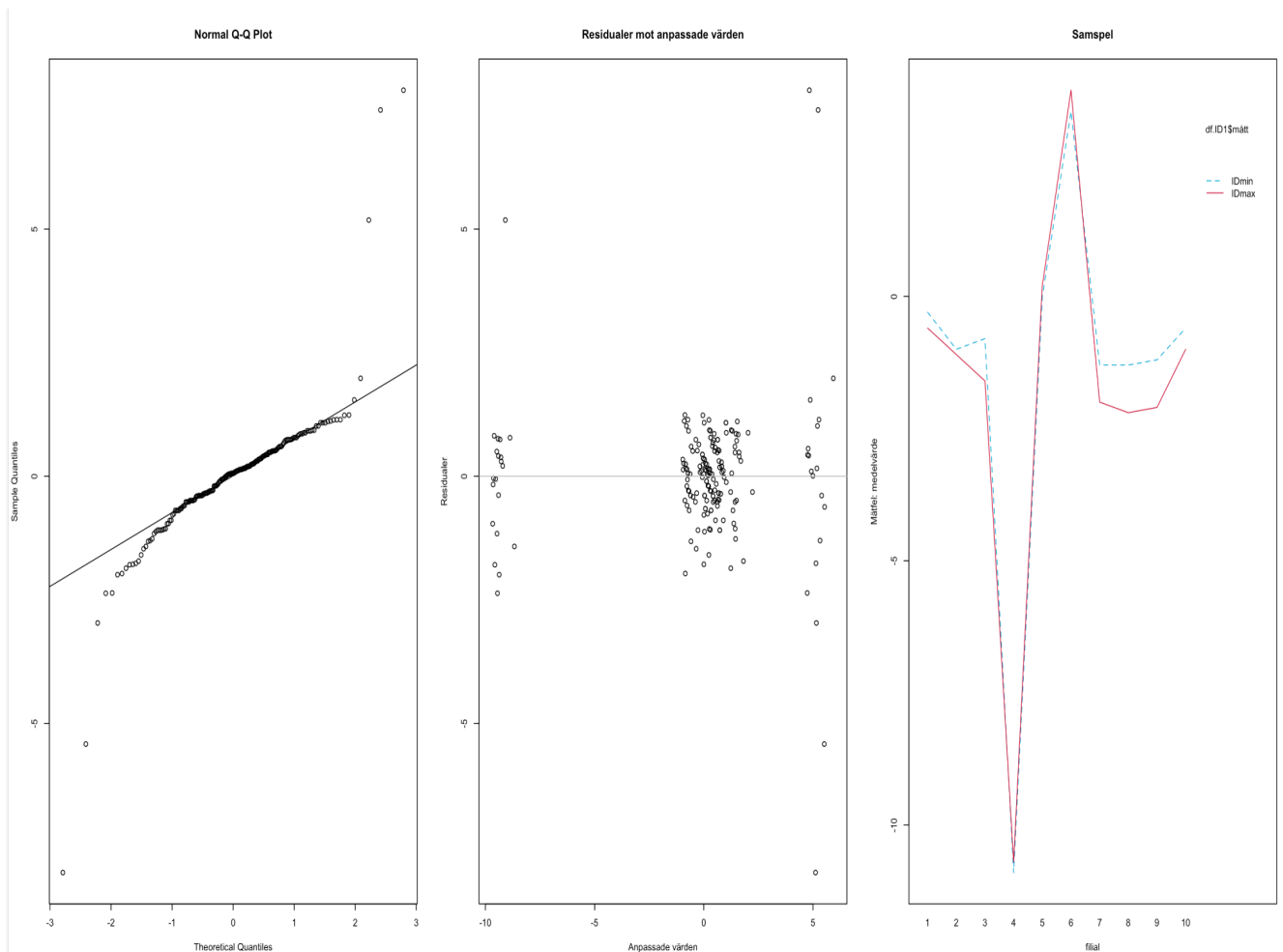
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	9	59.55	6.616		

Error: Within

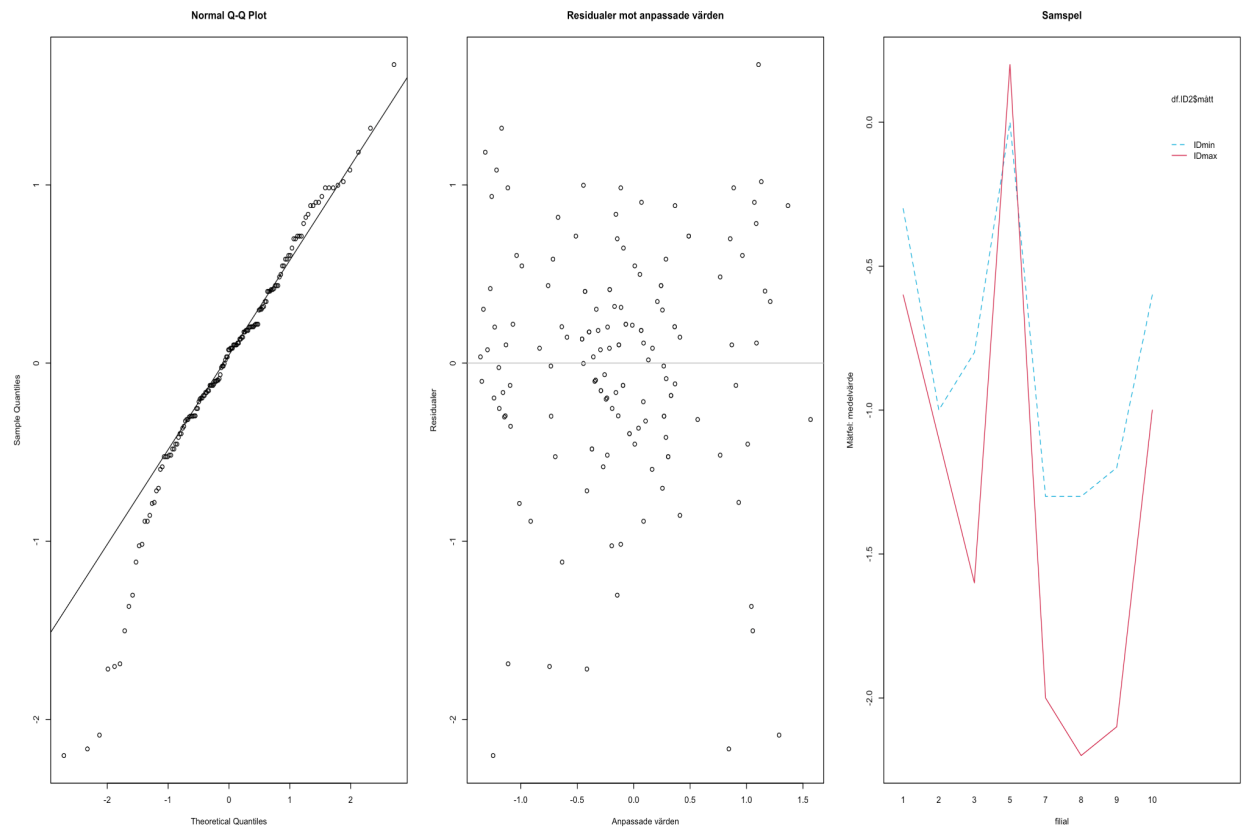
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
filial	9	2361.0	262.34	125.113	<2e-16 ***
mått	1	5.4	5.45	2.597	0.109
filial:mått	9	10.8	1.20	0.573	0.818
Residuals	171	358.6	2.10		

Signif. codes:

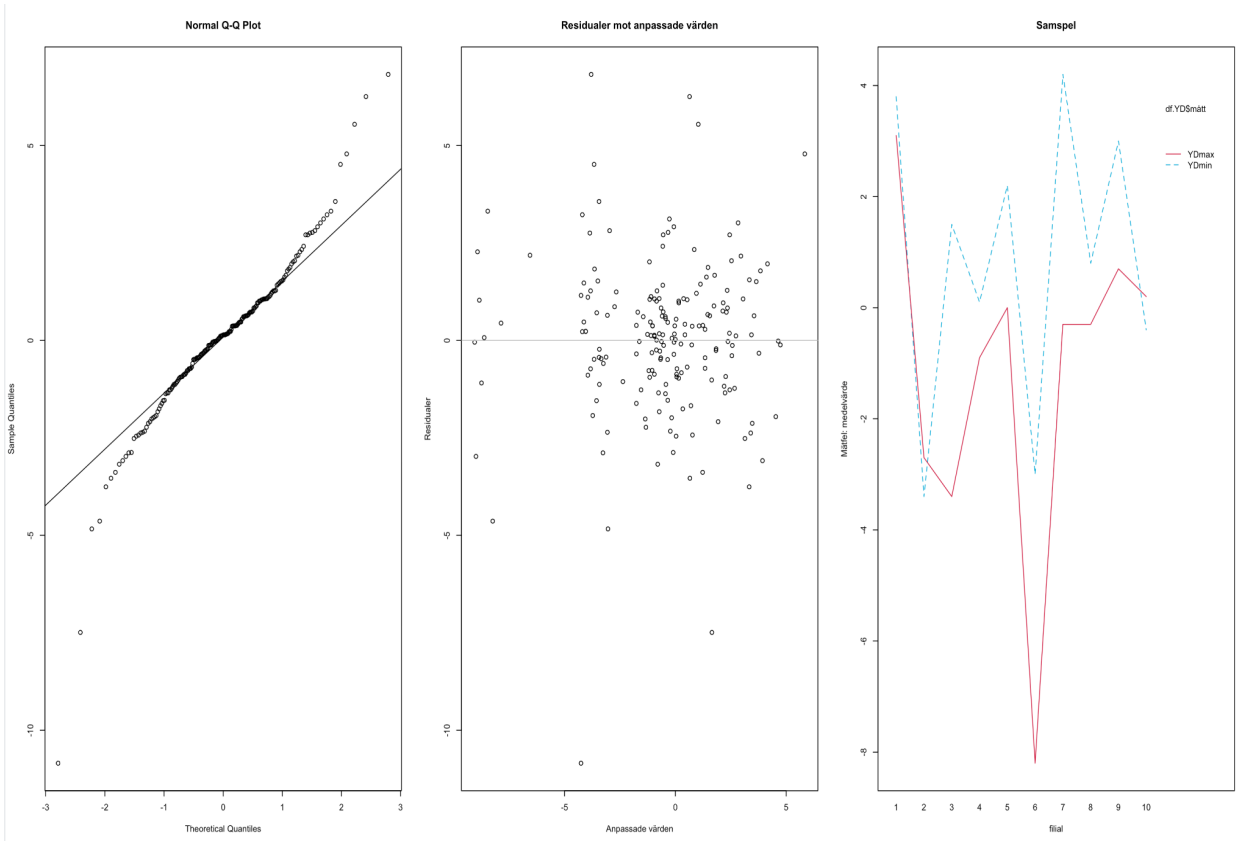
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Modell 1.3



Modell 1.4



Modell 1.5

